Mathematical Modelling of Engineering Problems
Vol. 10, No. 6, December, 2023, pp. 1987-1993
Journal homepage: http://iieta.org/journals/mmep

Machine Learning Based Detection of Fake Facebook Profiles in Afan Oromo Language

Kedir Lemma Arega[1], Mustafa K. Alasadi[2], Atyaf Jarullah Yaseen[3], Ayodeji Olalekan Salau[4,5*], Sepiribo Lucky Braide[6], Jeremiah Oluwatosin Bandele[7]

[1] Department of Information Technology, Ambo University, Ambo, Oromia 19, Ethiopia
[2] Department of Computer Science, University of Sumer, Al-Rifai 64005, Iraq
[3] Department of Computer Science, Thi-Qar University, Thi-Qar, Nasiriyah 0096442, Iraq
[4] Department of Electrical/Electronics and Computer Engineering, Afe Babalola University, Ado-Ekiti 360101, Nigeria
[5] Department of Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Tamil Nadu 600124, India
[6] Department of Electrical and Electronics Engineering, Rivers State University, Port Harcourt 5080, Nigeria
[7] Department of Electrical/Electronic Engineering, School of Science and Technology, Pan-Atlantic University, Ibeju-Lekki, Lagos 73688, Nigeria

Corresponding Author Email: ayodejisalau98@gmail.com

https://doi.org/10.18280/mmep.100608

Received: 10 April 2023
Revised: 15 July 2023
Accepted: 10 September 2023
Available online: 21 December 2023

Keywords:
Afan Oromo, machine learning, Facebook, fake accounts, feature selection, classification

ABSTRACT

The proliferation of social media platforms, including Facebook and Twitter, has significantly enhanced global communication and information exchange. Concomitantly, it has engendered the creation of counterfeit profiles, through which personal data is unscrupulously harvested and false information disseminated. This study presents a sophisticated ranking model designed to detect such fake Facebook profiles with utmost accuracy. An amalgamation of feature-based and graph-based techniques was utilized in this model, with the support vector machine (SVM) functioning as the primary tool for fraudulent account detection. The model was simulated using Python-based tools and the generated data subsequently subjected to in-depth analysis using the proposed strategy. Employing 5-fold cross-validation, the dataset was effectively trained and tested. Two successful categorization strategies were adopted to distinguish between legitimate and fraudulent social media accounts, thereby enhancing the detection of fake profiles. SVM was utilized to categorize the data, with a feature set technique applied for detection. The comparative efficiencies of neural networks (NN) and SVM were evaluated, demonstrating accuracies of 89% and 85% respectively. The proposed strategies exhibited superior accuracy of 93% when utilizing SVM and 89% with NN, outperforming existing methods. The approach primarily employed SVM, with a 5-fold cross-validation technique used to train and test the dataset. This validation method divided the dataset into five equal parts, enabling a comprehensive analysis of the results.

1. INTRODUCTION

Cushitic languages, including Afan Oromo, are widely spoken across nations such as Ethiopia, Kenya, Somalia, and Egypt. Social media platforms like Facebook, Instagram, Telegram, and WhatsApp are increasingly grappling with the issue of counterfeit accounts and deceptive content. With daily posts exceeding a trillion, traditional verification methods are rendered ineffective. Fake news carries the potential to not only mislead public opinion but also disrupt social order, undermine credibility, and engender a crisis in confidence.

This study introduces the application of machine learning to detect fake Facebook profiles in the Afan Oromo language. The escalating problem of fictitious profiles on social media platforms necessitates innovative solutions, prompting researchers to explore various methodologies. The objective of this research is the development of a machine learning model specifically tailored to the Afan Oromo language, predominantly spoken in Ethiopia and surrounding regions.

The intention of this paper is to bolster efforts to mitigate the presence of fake profiles, thereby enhancing online safety for users through the deployment of advanced algorithms and linguistic analysis. To surmount the associated challenges, a new dataset for the Afan Oromo language was generated. Ethiopians leverage social media platforms for myriad purposes ranging from political campaigning and religious discourse to information dissemination and participation in governmental policies [1].

Social media networks have soared in popularity in recent years, serving as vital conduits for connectivity, knowledge sharing, and personal online business operations [2, 3]. The exponential growth in internet usage has fueled the expansion of social networking sites, which have emerged as premier platforms for global information exchange. Users can express their viewpoints on a broad spectrum of subjects, with these web-based platforms offering an array of applications for communication, information, and updates [4].

However, the popularity and data abundance of these networks also provide fertile ground for the propagation of fake news by malicious actors, often through the creation of counterfeit accounts or unauthorized access to others' personal accounts. In 2019, Facebook reported the removal of approximately two billion fake accounts per quarter, with a subsequent disclosure of information regarding their employed machine-learning system [4, 5].

They identified two categories of fake accounts: “user-misclassified accounts,” essentially personal profiles for businesses that are easily converted to page views, and the more troubling “violating accounts.” These latter accounts, often used for scamming and spamming, violate the platform's terms of service and are notoriously challenging to control.

The subsequent sections of this paper are organized as follows: Section 2 provides a review of related works. Section 3 discusses the classification algorithms, while Section 4 describes the experimental framework and presents the experimental results. Finally, Section 5 delivers the conclusion and recommendations for future work.

2. RELATED WORK

Rahman et al. [6] presented a technique called FRAppE to detect malicious applications on Facebook. FRAppE alerts the user before installing the malicious application. It is a feature-based detection technique. This technique was developed using data collected from Facebook users online. The feature set identified helps in the detection of fake Facebook applications. To classify the data, a machine learning (ML) algorithm is used. In FRAppE, 5-fold cross validation was used to train and test data. FRAppE detects malicious applications with a 99.5% accuracy and a 95.9% true positive rate [6]. However, because it employs FRAppE alert, the study was not performed for data classification. As a result, it is ineffective in detecting Afan Oromo fake accounts [7].

Shekokar and Kansara [8] developed a social graph-based Sybil node finding method. By including user behavioral factors as latent transactions and friendship rejection, this approach overcomes the drawbacks of the earlier graph-based systems. The proposed system has two components: Sybil node identification (SNI) and Sybil node identification using behavioural analysis (SNI-B). Traditional Sybil detection was used in the SNI approach, which is an extension of SNI. The outcomes of both approaches were compared. Compared to SNI, SNI-B offers greater accuracy, precision, and recall. SNI's accuracy is 77% for SNI-B and 92% for SNI. SNI's accuracy is 75% for SNI-B and 80% for SNI. Similarly, SNI recall is 60% and SNI-B recall is 80%. Some topics are well-explained in this paper, but the social graph is the only factor that matters. As a result, the authors did not employ the fraudulent account detection method for Afan Oromo. ElAzab [9] presented classification techniques to detect fake accounts on twitter. To detect the fake accounts the authors used feature-based approach. The minimum weighted feature set is used. In this approach, the behavior of the user was identified and it was observed that the real user behaves differently than the fake users. This behavior was used to identify the fake accounts. The accuracy of all techniques is provided. The gain measure is used to assign the weights to the feature set. To train and test the algorithms, 5-fold cross validation was applied and SVM gives best accuracy results to detect the fake accounts.

The study in Meligy et al. [10] introduced a method called fake profile recognizer for identifying phony social networking profiles. This method is based on deterministic finite automata and regular expressions. The profiles are verified using a regular expression, and deterministic automata reliably identify the individuals. This method is used with data from Twitter, Google Plus, and Facebook. Facebook, Twitter, and Google+data sets' accuracy rates are 89.73%, 76.94%, and 81.9%, respectively. The precision for Twitter is 81.81%, 77.41% for Google+, and 88.9% for Facebook. Facebook has a false positive rate of 11.66, Google+at 26.10, and Twitter at 20.86%. False negative rates for Twitter are 18.20%, Google +22.60%, and Facebook are each 11.04% [11].

Mohammadrezaei et al. [12] presented an effective method using machine learning methods. The classifier was trained by 10-fold cross-validation; thereafter efficiency metrics were calculated. First, the cross-validation technique was defined, followed by the basic metrics and evaluation of classifier performance. The data contains information about the existing relationships between the nodes, which is then used to generate the graph's adjacency matrix and calculate measures of similarity between nodes. The PCA technique is then used to extract new features. The SMOTE was then used to generate artificial data. Data distribution is altered by using SMOTE. It means that the 99% normal users and 1% fake users have been changed to 75% normal and 25% fake, and the balanced data has been sent to the next step. However, SMOTE's 10-fold cross-validation and the percentage given for a fake account user are small. It is also ineffective for detecting Afan Oromo fake accounts.

2.1 Overview of social media

On-line social networks are popular channels to stay in contact thereby enabling people to communicate, and share their everyday activities, photos, and status. Social networking sites like Facebook are very popular among people. Social media is a phenomenon which links people to each other in so many ways. Rivera, a social media service focuses on building online communities of people who share interests and/or activities, or who are interested in exploring the interest and activities of others [7]. People create networks of friends and associates there. It is used to forge and establish connections amongst various people. Social media can play a significant role in networking and collaboration. Social networking websites are frequently employed as a form of inter-person communication. Social networking site users can exchange information and activities of their day-to-day life, thus drawing a lot of individuals to these sites [13].

2.2 Social media usage in Ethiopia

The development of a country depends on technology implemented based on political, culture, religion, economic, and social background. In sub-Saharan countries, social media platforms have increased usage, leading to misuse. Traditional media platforms have limited regulatory oversight, leaving platform providers to remove fake content. In 2018, Ethiopia had the world's second-highest relative social media growth, translating to the 16th-highest absolute growth globally and the third-highest absolute growth in Africa, trailing only the continent's two economic powerhouses, South Africa and Nigeria [14]. The 2017 report "feasibility Study-Strengthening Free, Independent and Professional Journalism in Ethiopia" should be read in conjunction with social media usage in Ethiopia which shows that social media is a risky underdog [14-16]. In just a few years, the social media has fundamentally altered Ethiopia's political environment, serving not only as a tool for populist activism but also for the spread of misinformation, hate speech, and rumors.

In the Afan Oromo language, this study looks into the problem of bogus Facebook profiles. Previous studies on the subject have concentrated on cloning, social network analysis, online fake profile identification using these techniques, as well as FakeBr, a framework for identifying phony identities in online social networks. The work attempts to address the particular problem of identifying false profiles in the Afan Oromo language, which necessitates the creation of language-specific methods and models.

The authors have developed various methods to detect fake accounts on Facebook, including FRAppE, a feature-based detection technique, and SNI-B, a social graph-based Sybil node finding method. FRAppE detects malicious applications with 99.5% accuracy and a 95.9% true positive rate, but it is ineffective for detecting Afan Oromo fake accounts. El Azab et al. introduced a method called fake profile recognizer, which uses deterministic finite automata and regular expressions to identify phony profiles. Meligy's fake profile recognizer has 89.73% accuracy rates, 76.94% precision, and 81.9% accuracy for Twitter, Google Plus, and Facebook, with a false positive rate of 11.66, Google+at 26.10, and Twitter at 20.86%. Mohammadrezaei et al. [12] method uses machine learning methods, with a 10-fold cross-validation and small percentage given for a fake account user. However, SMOTE's 10-fold cross-validation and small percentage given for a fake account user make it ineffective for detecting Afan Oromo fake accounts.

2.3 Contributions of the paper

In this paper, the fake Facebook accounts were detected by using machine learning techniques. Two techniques were used to classify the authentic and fake accounts. Feature set approach is used for the detection. Two machine learning methods were evaluated, namely: NN and SVM. These techniques accept the acquired random data and provide the most accurate results.

3. METHODOLOGY

This section presents the method employed for data collection and analysis. In addition, the statistical method employed for analyzing the dataset was presented. Furthermore, the performance metrics employed in the evaluation of the performance of the employed machine learning algorithms are discussed. Data collection, feature extraction, labeling, splitting the dataset, preprocessing, model selection, feature engineering, model training, evaluation, tuning and validation, testing, deployment, and monitoring are just a few of the steps in the methodology for identifying fake Facebook profiles in Afan Oromo.

3.1 Dataset acquisition

A corpus of counterfeit reports from various accounts was used in an experiment using Afan Oromo's Facebook accounts. The corpus was put together by specialists in the field after going through tokenization and normalization processes and the process of obtaining data entails compiling a sizable dataset of real and fictitious Facebook accounts in Afan Oromo, which can be acquired by web scraping or open Facebook APIs. Acquiring the Facebook dataset is the first step in identifying a phony account on social media platforms. Facebook's data was acquired for this work using a surveying technique. We gathered the Facebook dataset from Facebook users using the survey approach. To accomplish this, we developed a Google form. Different types of inquiries on the Google form enable us to distinguish between real and phony accounts to a certain accuracy. Online data was gathered through Google forms and manually completed forms by Facebook users. We gathered information from 1654 Facebook accounts. From the Facebook dataset that has been compiled, we extracted different features. The 16 elements in these feature sets assisted in precisely classifying the data.

1. Number of Facebook friends.
2. Number of photos shared on Facebook.
3. Number of status/news shared per month.
4. Number of groups joined.
5. Number of likes made per day.
6. Number of days since the profile has been updated.
7. Year in which the user joined Facebook.
8. Number of pages liked.
9. Number of posts liked by a Facebook friend.
10. Account user has a profile photo.
11. Account as a cover photo.
12. Frequently used hashtags in posts.
13. Account is logged in a location.
14. Account is logged in using time zone.
15. Number of Facebook friend that tagged the user.

3.2 Dataset preparation

By using the survey method, we created a Facebook dataset from Facebook users. For this purpose, we developed a Google form. The google form consists of various types of questions, which help us to accurately classify the data into real accounts and fake accounts. The dataset was collected online by using by Facebook users manually filling the google forms. We collected data of 1654 Facebook accounts. Thereafter, data was gathered, prepared using a procedure that includes gathering, cleaning, filtering, and consolidating the data into one file or data table [7]. Tools for implementing this approach include MS Excel. The data cleaning and filtering process was largely used for the annotation of user activities and timelines in the dataset at the following stages, which is followed by the creation of a training model for Afan Oromo false account detection. To get the dataset ready for annotation, the following steps were performed:

Eradicate all user activity and information that isn't Afan Oromo and doesn't involve text.

i. Eliminating all whitespace, blank values, and nulls.
ii. Filtering out with words that represent the vocabulary of phony accounts.
iii. In order to ensure that each text in a dataset is unique, each page's data is combined into a single dataset.
iv. Preserve the context of each text in a dataset for the annotation processes.

All of the aforementioned dataset preparation steps takes into account the nature and behaviours of the Afan Oromo language. The keywords are collected from several social media user pages that were well-known for employing fake account names to aid the filtering process.

3.2.1 Dataset filtration

We used a randomized filtration technique to filter the acquired dataset. Feature extraction is often carried out when the original raw data is significantly different and cannot be used for machine learning modeling. Feature extraction is the process of creating new, more precise features from raw data that capture the majority of its pertinent information. When working on real-world ML challenges, data is typically obtained in CSV format, thus the pertinent features from the raw data must be extracted. One of the many feature extraction methods that we used is the TF-IDF vectorizer. The position of the accounts in the dataset is altered randomly through randomization. The dataset's false values are likewise filtered using this method, and the false values are then replaced with the average of its upper and lower column values. To accurately categorize the dataset, filtering was applied and it was observed that the dataset was accurately classified by the classification method and contained few incorrect or null values.

3.2.2 Clustering of the dataset

K-means clustering and PCA were chosen for their strengths and suitability in identifying patterns and groupings in data. K-means clustering partitions data into K clusters based on similarity, providing insights into customer behavior and preferences. PCA, a dimensionality reduction technique, transforms high-dimensional datasets into lower-dimensional spaces, identifying important features, and visualizing data. Both algorithms can be tuned to achieve optimal results. The dataset was subjected to the clustering algorithm in addition to filtration. The K-means clustering approach was used to group the dataset into clusters. For the dataset, there are two clusters, Cluster A and Cluster B. Cluster A is made up of fake account data and actual account data in Cluster B. The clustering algorithm detects numerous phony accounts, thus improving accuracy and simplifying the process. The clustering approach was used to locate the fake accounts. Because K-mediod is more resistant to outliers than K-mean, it is a better clustering algorithm.

3.3 Feature selection

In the feature selection stage, the feature set is subjected to the feature selection technique. Finding pertinent features, such as account age, friend count, profile picture quality, posting frequency, likes, comments, and shares, is the process of feature extraction. The dataset is divided into training, validation, and testing sets to assess the model's performance. Labeling aids in distinguishing between real and fraudulent profiles. The eigenvalues and eigenvectors are calculated as part of principal component analysis, and they are ranked from the highest to the lowest. Additionally, the associated traits are combined and given a value by principal component analysis. Using the feature selection strategy, the accuracy was increased while using fewer features. The lowest weight zero can exclude the characteristics when employing the feature selection technique. Table 1 presents the ranking attributes/features selected. In our proposed study, we trained and tested the data using the 5-fold cross validation method. The 5-fold cross validation technique is used to assess and test the classification algorithm. This involves breaking up the training set into five smaller sets and analyzing the outcomes. The dataset is divided into 5 sections using the 5-cross fold validation method. The dataset is divided into four training sections and one testing part. With five separate training and testing datasets, the best training and testing process is repeated.

The model and hyperparameters are adjusted during tuning and validation based on the findings of the evaluation. The independent labeled testing dataset is tested to evaluate its generalizability and overall accuracy on unobserved data. The model must also be periodically updated and retrained to ensure that it is still effective in identifying phony Facebook profiles in Afan Oromo. Deployment and monitoring activities complete the process.

Table 1. Ranking attribute/feature selection

Ranking	Attribute
0.6421	Number of Facebook friends, Number of friends you tagged, likes per day, Account has a cover photo.
0.5327	Number of groups you joined, pages liked, number of photos, Number of Facebook user liked the post, frequently used hashtags in posts.
0.4556	In which year joined Facebook, Number of new feed shared, pages liked, and Number of groups joined, number of photos.
0.397	Number of days since updated the profile, in which year joined Facebook, frequently used hashtags in posts, Number of new feed shared, logged in an account.
0.3417	Number of Facebook friends tagged the user, Facebook friends tagged, logged in an account using iPhone, number of new feed shared, number of posts liked by Facebook users, Number of friends you tagged.
0.2465	Logged in your account using phone, number of photos, logged in account using Android phone, number of days since updated the profile, in which year joined Facebook.
0.2043	Number of new feed shared, number of photos, logged in account using Android phone, in which year joined Facebook, frequently used hashtags in posts.
0.178	Number of photos, number of new feeds shared; logged in your account using iPhone, number of Facebook user liked your post, Number of groups joined.
0.1343	Number of Facebook user liked your post, Number of groups joined, number of new feeds shared, pages liked, logged in account using Phone.
0.1046	Logged in account using iPhone, frequently used hashtags in posts, account is connected with Instagram, Number of days since updated the profile, in which year joined Facebook.
0.0879	Account has a profile photo, logged in account using Android phone, Account is connected with other, number of new feed shared, Account has a cover photo.
0.057	Account has a cover photo, logged in account using Android phone, pages liked.
0.0385	Number of likes per day, Number of Facebook friends, Number of friends tagged.

4. RESULTS AND DISCUSSION

This section shows how the suggested techniques work and also presents how the methods performed for Afan Oromo language fake profile detection. The best machine learning algorithm and feature extraction technique are selected for the final model comparison in this study using an experimental design. The best model is then used to develop a prototype for fake account detection. The fraudulent accounts were identified on Facebook utilizing the machine learning techniques. The classification of authentic and fake accounts is done using the two most effective methods. The detection is done using a feature set technique. The feature set that affects the detection of fake accounts is identified. SVM will be used in the suggested study to locate the most accurate outcomes. These methods use random data and get the most precise findings.

4.1 Comparison with existing technique

The comparison between existing and the new proposed hybrid technique is shown with the various parameters. These parameters show that the proposed hybrid technique performs better than existing techniques. The comparison of NN and SVM in proposed technique is also compared. The parameters used to compare three results of both techniques are Accuracy, Precision, Recall, F-measure and Execution time. There are following parameters are used to compare proposed hybrid technique with an existing technique.

4.1.1 Accuracy

Accuracy is the proportion of bogus accounts that were accurately recognized. It can be calculated as the proportion of accounts that were correctly categorized to all accounts. For the best technique, it ought to be at its peak. The proposed strategy is more accurate than the current method. SVM and current NN both have accuracy levels of 85%. With the suggested hybrid approach, NN and SVM accuracy is 89% and 93%, respectively.

4.1.2 Precision

It evaluates how many accounts have correctly classified a fake among those all that are classified as fake. It can be measured by a number of accounts that are correctly classified as fake to the total number of accounts classified as fake. Figure 1 shows the comparison of an existing technique and proposed technique based on precision also. The precision should be high. The precision of existing NN and SVM is 0.88 and 0.91 respectively. The precision of NN and SVM in proposed technique is 0.30 and 0.84 respectively.

4.1.3 Recall

The percentage of accurate documents chosen in class that genuinely belong to the class, as well as recall, should be high. Figure 1 compares an intended technique based on recall with an existing technique. Existing NN has a recall of 0.55, whereas existing SVM has a recall of 0.97. The results of the suggested hybrid approach are respectively 0.74 and 0.62. The graph demonstrates that the hybrid technique recall is higher than the current technique. The true positive rate and false negative rate are used to compute the recall.

4.1.4 F-Measure

It is defined as the weighted harmonic mean of Precision and Recall. It is also known as F-score. F-measure calculated from Precision and recall. F-measure also should be high. The Figure 1 also shows the comparison between f-measure of an existing technique and proposed technique. F-Measure of NN and SVM in existing technique is 0.67 and 0.94 respectively. The f-measure of NN and SVM in supposing the hybrid technique is 0.42 and 0.71 respectively, which is higher than the existing technique. The Precision of SVM in proposed technique is very much higher than the precision of existing technique. The graph shows that the f-measure of proposed hybrid technique is higher than the existing technique. The evaluated result and experiment classification model by using the 5-fold cross-validation method, achieving an average result as presented in Figure 2 and Table 2 for the fake account classification performance using SVM, and SVM and NN in Table 3, while Figure 3 shows the results of the CNN model.

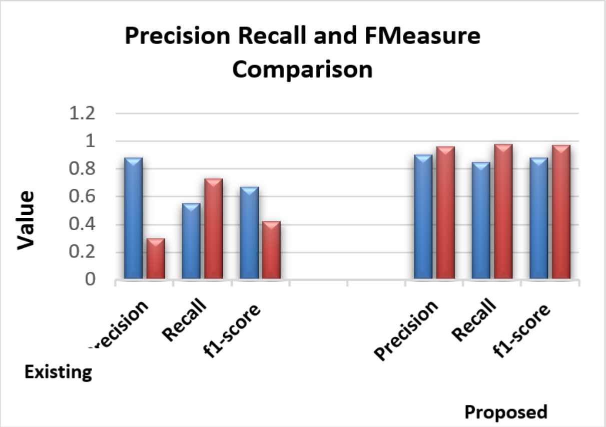

Figure 1. Comparison based on Recall, Precision, and F-measure

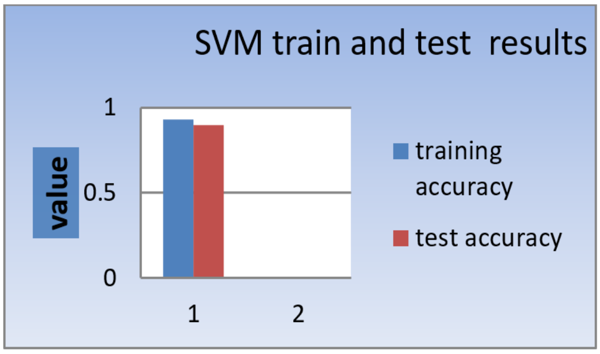

Figure 2. Train and test accuracy for SVM

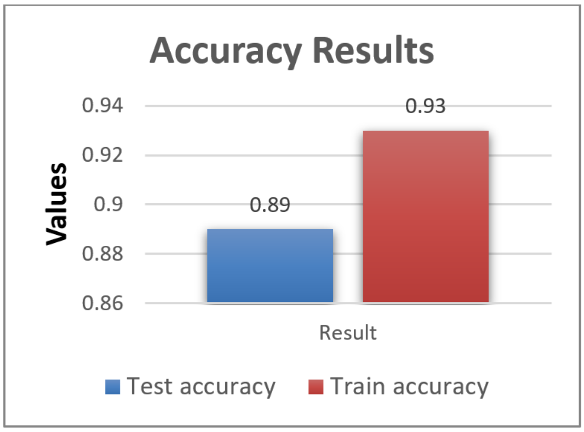

Figure 3. Results of test and train accuracy

Table 2. Accuracy of SVM

	Precision	Recall	F1-Score	Support
0.0	0.90	0.97	0.94	262
1.0	0.85	0.60	0.71	68
Average	0.92	0.92	0.90	330

Table 3. Results of SVM and NN

	Precision	Recall	F1-Score	Support
0.0	0.96	0.98	0.97	251
1.0	0.90	0.85	0.88	66
Average	0.95	0.95	0.95	

i. Recall

Recall answers what proportion of actual positives was correctly identified. Recall calculated using the Eq. (1).

$$\text{Recall}=\frac{\text{TP}}{\text{TP+FN}} \quad (1)$$

where, T is the number of correct predictions and N is the total number of predictions (correct and incorrect) made by a given classification model.

ii. Accuracy

It is a measure of the correct number of predictions to the total number of predictions in the data. Accuracy is highly reliable for balanced datasets. Accuracy shows the classification problem correct prediction value and calculates as the total number of the model correct prediction divide by all number of data instances used for the model accuracy calculated using the Eq. (2).

$$\text{Accuracy}=\frac{\text{TN+TP}}{\text{TP+TN+FP+FN}} \quad (2)$$

5. CONCLUSION

This research investigates fake account detection for Afan Oromo language on social media to tackle fake crime and contribute to better detection systems and dataset development. The research "Identifying Fake Facebook Profiles in Afan Oromo Using Machine Learning" tries to address the problem of phony profiles in the Afan Oromo language. The authors suggest a machine learning-based method that successfully detects fraudulent profiles with high rates of accuracy. By examining characteristics like friends, posts, comments, and profile images, the algorithm successfully distinguishes between authentic and phony profiles. As a useful tool for managing the rising popularity of social media platforms, the suggested strategy shows promise in addressing fraudulent profiles in Afan Oromo on Facebook. As different languages may have unique qualities or patterns that might be used for identifying purposes, the study stresses the significance of language-specific approaches in combating bogus profiles. The paper shows the viability and efficacy of applying machine learning to detect fraudulent Facebook profiles in Afan Oromo, and it has the potential to be expanded further and used as a useful tool to reduce the dissemination of incorrect information and safeguard user privacy on social media platforms.

This paper presents two machine learning methods for detecting fake accounts in Afan Oromo language on social networks like Facebook. The hybrid approach uses neural networks (NN) and SVM classifiers, with K-mediod clustering to increase precision and reduce time complexity. A real-time dataset was acquired, and the dataset was filtered using randomization techniques. The classification technique was applied to detect multiple fake accounts simultaneously. Principal component analysis (PCA) was used to rank feature sets, resulting in better accuracy than other algorithms. The 5-fold cross validation method was applied for training and testing the dataset. The accuracy of the proposed work with NN and SVM was 89% and 85%, respectively, while the precision was 93%.

REFERENCES

[1] Guta, L. (2019). Social network hate speech detection for afaan oromoo language, p. 8.

[2] Arega, K.L. (2020). Social media fake account detection for amharic language using machine learning. GSJ, 604-614.

[3] Demilie, W.B., Salau, A.O. (2022). Detection of fake news and hate speech for Ethiopian languages: A systematic review of the approaches. Journal of Big Data, 9(1): 66. https://doi.org/10.1186/s40537-022-00619-x

[4] Appel, G., Grewal, L., Hadi, R., Stephen, A.T. (2020). The future of social media in marketing. Journal of the Academy of Marketing Science, 48(1): 79-95. https://doi.org/10.1007/s11747-019-00695-1

[5] Kaba, G.D., Olalekan, S.A., Asrat, G. (2022). Afaan oromo language fake news detection in social media using convolutional neural network and long short term memory. Journal of Electrical and Electronics Engineering, 15(2): 37-42.

[6] Rahman, S., Huang, T.K., Madhyastha, H.V., Faloutsos, M. (2015). Detecting malicious facebook applications. IEEE/ACM Transactions on Networking, 24(2): 773-787. https://doi.org/10.1109/TNET.2014.2385831

[7] Bhambar, S., Kanchan, Khairnar Yogita Nikam. (2022). Detecting fake accounts on social media using. International Research Journal of Modernization in Engineering Technology and Science, 1-4.

[8] Shekokar, N.M., Kansara, K.B. (2016). Security against sybil attack in social network. In 2016 International Conference on Information Communication and Embedded Systems (ICICES), Chennai, India, pp. 1-5. https://doi.org/10.1109/ICICES.2016.7518887

[9] ElAzab, A. (2016). Fake accounts detection in twitter based on minimum weighted feature set. World Academy of Science Engineering and Technology, 10(1): 13-18.

[10] Meligy, A.M., Ibrahim, H.M., Torky, M.F. (2017). Identity verification mechanism for detecting fake profiles in online social networks. I.J. Computer Network and Information Security (IJCNIS), 9(1): 31-39. https://doi.org/10.5815/ijcnis.2014.01.01

[11] Altay, S., Berriche, M., Acerbi, A. (2023). Misinformation on misinformation: Conceptual and methodological challenges. Social Media+Society, 9(1): 20563051221150412. https://doi.org/10.1177/20563051221150412

[12] Mohammadrezaei, M., Shiri, M.E., Rahmani, A.M. (2018). Identifying fake accounts on social networks based on graph analysis and classification algorithms. Security and Communication Networks, 2018. https://doi.org/10.1155/2018/5923156

[13] Virdi, P. (2017). Fake accounts detection in Facebook using machine learning techniques. Phagwara, Punjab, pp. 1-61.

[14] Christer, N. (2019). Social media and journalism in Ethiopia. Linnaeus University Stockholm: Fojo Media Institute.

[15] Pulluri, S.R., Gyani, J., Gugulothu, N. (2017). A comprehensive model for detecting fake profiles in online social networks. International Journal of Advanced Research in Computer and Communication Engineering, 6(6).

[16] Ababu, T.M., Woldeyohannis, M.M. (2022). Afaan Oromo hate speech detection and classification on social media. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, Marseille, France, pp. 6612-6619. https://aclanthology.org/2022.lrec-1.712